# REDUCING RESIDUAL NOISE USING THE INTELLIGENT HEARING SYSTEM (IHS) AND THE VIVOSONIC INTEGRITY AUDITORY BRAINSTEM RESPONSE DEVICES

by

Julie Thein



Audiology Doctoral Project Submitted to the Faculty of the

DEPARTMENT OF SPEECH, LANGUAGE, AND HEARING SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF AUDIOLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Audiology Doctoral Project Committee, we certify that we have read the project prepared by Julie Thein, titled "Reducing Residual Noise using the Intelligent Hearing System (IHS) and the Vivosonic Integrity Auditory Brainstem Response Devices" and recommend that it be accepted as fulfilling the Audiology Doctoral Project requirement for the Degree of Doctor of Audiology.

Linda Norrix Date: 4/8/19
Linda Norrix, Ph.D.

Date: 4/8/19
David Velenovsky, Ph.D.

Date: 4/8/19
Huanping Dai, Ph.D.

Date: 4/8/19
Michael Hartley, Ph.D.

Final approval and acceptance of this project is contingent upon the candidate's submission of the final copies to the Graduate College.

I hereby certify that I have read this Audiology Doctoral Project prepared under my direction and recommend that it be accepted as fulfilling the project requirement.

Linda Norrix Date: 4/8/19
Audiology Doctoral Project Director: Linda Norrix, Ph.D.

Date: 4/8/19
Audiology Doctoral Project Director: David Velenovsky, Ph.D.

## ACKNOWLEDGEMENTS

Completing this research over the past year would not have been possible without the support and guidance of the following individuals. I want to thank each one of them for believing in me when I did not believe in myself, for providing guidance, and for supporting me in this research.

Linda Norrix, Ph.D.

David Velenovsky, Ph.D.

Huanping Dai, Ph.D.

Michael Hartley, Ph.D.

Mark Borgstrom, Ph.D.

Barbara Cone, Ph.D.

Joseph Vasey, B.S.

# DEDICATION

After spending a year working on this research, I would like to dedicate my work to my parents, Cathee and Randy, and to my boyfriend, Brian. You picked me up when I was at my lowest points and provided the support and confidence that I needed to raise me higher than I ever thought possible. Without all of you I would not have been able to finish this project and for that I thank you.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

ABSTRACT

**Purpose:** The purpose of this study was to determine the effectiveness and efficiency of the Intelligent Hearing Systems (IHS) (utilizing artifact rejection) and the Vivosonic Integrity (VIVO) (utilizing Kalman weighted averaging, the Amplitrode, and Bluetooth) in reducing residual noise.

**Method:** Simultaneous ABR recordings were collected for 16 adults during both relaxed and active motor states. Residual noise (RN) measures obtained using artifact rejection as implemented on the IHS were compared to those obtained using Kalman weighted averaging, the Amplitrode, and Bluetooth technologies as implemented on the VIVO.

**Results:** Average RN levels obtained by VIVO were lower than those obtained by the IHS in all conditions. With use of Kalman weighted averaging, the Amplitrode, and Bluetooth, 3 minutes of averaging for the relaxed condition and 6 minutes of averaging for the active conditions were needed in order to meet a criterion RN level of 0.025μV.

**Conclusions:** The use of Kalman weighted averaging provides an advantage in both effectiveness and efficiency over traditional averaging and artifact rejection in reducing residual noise levels.

# INTRODUCTION

The auditory brainstem response (ABR) is a physiologic measure that is used clinically to estimate hearing thresholds in adults who are unable or unwilling to respond using behavioral test techniques, to assess the neurological integrity of the auditory nervous system pathways, and most commonly, to estimate behavioral hearing sensitivity in newborn and infant populations (Lightfoot & Stevens, 2014). The ABR is an auditory evoked potential that includes 5-7 peaks that arise from the auditory neurons (Picton, 2011). Wave V is the most important of the peaks when estimating hearing thresholds as it is the most robust and its presence is highly correlated with behavioral hearing thresholds. The ABR is obtained by placing electrodes on the surface of the scalp, playing sounds (e.g., clicks, tonebursts, chirps) to the ear(s) and recording the ongoing electrical activity, over a period of ~10-15 ms, in response to the sounds.

When used to estimate hearing sensitivity, ABR wave V responses must be detected visually on a display screen. The threshold wave V response is defined as the lowest stimulus level that results in a response, where the response is not visible when the stimulus level is reduced by 5 or 10 dB. The threshold wave V response has a small amplitude that is much lower than the physiological background noise that is also recorded by the electrodes. In order to visually-observe a wave V response, the amplitude of the noise must be lower than the response amplitude. Further, on runs below threshold (to demonstrate that no response is present) the noise must also be low to confidently say that an ABR is not present; that is, it is unlikely that a small response is being obscured by the background electrophysiological noise. To reduce the amplitude of the noise, averaging is used. If noise is random and stationary, it will be reduced in the final averaged waveform by the square root of the number of sweeps (Picton, 2011). Thus, the greater the number of sweeps, the lower the residual noise (RN) in the averaged waveform. In addition to averaging, several techniques have been developed to reduce RN. Two such techniques are artifact rejection combined with traditional averaging and weighted averaging, in lieu of traditional averaging.

- Artifact rejection (AR): AR can be used to eliminate sweeps with amplitudes higher than the artifact rejection level criterion set by the audiologist. For a relaxed and sleeping infant, the recommended AR level criterion is between ±3 and ±10 μV (NHSP, 2013).
- Weighted averaging such as Bayesian weighting (Elberling and Wahlgreen, 1985; Don and Elberling, 1994) or Kalman weighting (Kay, 1993; Maybeck, 1979; Upp, 2002) can reduce noise in ABR recordings. Weighted averaging allows all sweeps into the average but will give more weight to the low-noise sweeps, or blocks of sweeps, and less weight to the high noise sweeps or blocks of high noise sweeps.

A few studies have compared the electrophysiological recordings obtained using traditional averaging with AR versus weighted averaging (Cone & Norrix, 2015; Don & Elberling, 1994; Sanchez & Gans, 2006). In 1994, Don and Elberling explored the effectiveness of AR and Bayesian-weighted averaging in reducing residual noise in eight participants. Residual noise was calculated using the variance approach described by Elberling and Don (1984), which calculates RN as the variance between individual sweeps at a single point. The authors did not report how participants were instructed to behave (quiet/relaxed or active) during the ABR recordings but noted that the participants had varying levels of electrophysiologic noise during the recordings. Click stimuli at 10 near-threshold stimulus levels ranging from 30 to 48 dB peak-to-peak equivalent sound pressure level were used to obtain 10,000 sweeps for each stimulus level. Sweeps were then submitted to off-line data analyses that included traditional averaging with seven levels of AR (ranging from ±2.5μV to ±10μV), and weighted averaging. Traditional averaging plus AR was performed by rejecting any digitized value between 1 and 11 ms poststimulus onset that exceeded the criterion rejection level. The remaining sweeps were then summed and divided by the total number of sweeps. Weighted averaging using Bayesian estimation principles was performed by weighting blocks of sweeps inversely to the level of noise calculated for each block. By examining the estimated RN in the averaged waveform using traditional averaging plus AR, and Bayesian weighting for each participant, the authors made the following observations and conclusions:

1. Background noise levels of participants oftentimes do not remain stationary over the course of testing.
2. Although with stringent rejection levels there will always be less noise allowed into the average, more sweeps will need to be averaged to reduce noise (noise is reduced by the square root of the number of sweeps). In some cases, averaging more sweeps that are noisy may result in a lower RN than averaging fewer sweeps with lower-amplitude noise.
3. Bayesian weighting reduced RN to a greater extent than did any artifact rejection level.
4. Bayesian weighting controls for episodic noise as the block in which the noise occurs will be weighted less than less noisy blocks. However, Bayesian weighting would be of no advantage to averaging if noise was constant as all blocks would be weighted similarly.
5. In comparison to traditional averaging and the use of AR, Bayesian weighting, which minimizes the destructive effects of episodic noise, improves the quality of objective statistical techniques and results in more accurate and efficient ABR testing.

Sanchez and Gans (2006) expanded on the Don and Elberling (1994) study by more systematically examining how AR and Bayesian-weighted averaging impacts the ABR wave V amplitudes and the RN in the averaged waveforms. They examined 20 adults who participated in both a quiet (relaxed) condition and an active (periodic head and mouth movement) motor condition. The stimuli used to generate the ABRs were rarefaction clicks presented at a rate of 25.1/s at 104 and 74-dB peak-to-peak equivalent sound pressure levels. These levels corresponded to 60 and 30 dB nHL, respectively. Similar to Don and Elberling (1994), Sanchez and Gans recorded ABRs using a large number of sweeps (i.e., 16,384) and submitted the sweeps to off-line analyses that included Bayesian-weighted averaging and AR. Bayesian weighting was performed using a 256 sweep block size. Weighting was performed for both 16 blocks (4096 sweeps) and 64 blocks (16,384 sweeps). For AR, two rejection levels were examined, $AR_{10}$ and $AR_{EN}$ (equal noise). $AR_{10}$ used an artifact rejection level of ±10 μV. To calculate $AR_{EN}$ all 16,384 sweeps were used and the rejection level was systematically reduced in 1 μV steps until the RN was equal to the RN calculated for Bayesian-weighted averaging using 16 blocks or 4096 sweeps. Custom software was used to determine wave V peak to trough amplitudes and noise RMS values. Unlike Elberling and Don (1984) who used the variance of a single point to estimate RN, Sanchez and Gans estimated RN by calculating RMS levels for the first 4,096 accepted sweeps after theoretically eliminating any evoked responses. They eliminated evoked responses by storing consecutive sweeps in alternate buffers and subtracting the waveforms in the two buffers. By using offline analyses of AR and Bayesian weighting and systematically varying participants' activity level during averaging, Sanchez and Gans (2006) found the following regarding wave V amplitude:

- For the quiet conditions, wave V amplitude did not vary as a function of noise reduction technique.

- In the active conditions, wave V amplitude was significantly larger using Bayesian weighting than either AR strategy. The authors concluded that nonstationary noise will influence whether the noise reduction techniques are effective in extracting the response from the noise.

Several findings were noted when examining the RN RMS values. In the quiet conditions, RN did not vary as a function of noise reduction technique. However, in the periodic motor activity condition, $AR_{10}$ (4096 sweeps accepted in the final average) resulted in a significantly lower level of RMS calculated noise than did Bayesian weighting (4096 total sweeps with individual blocks weighted inversely to the amount of noise in that block) or $AR_{EN}$ (AR level chosen to closely match the RN obtained for Bayesian weighting). In contrast, the RMS noise levels were comparable when $AR_{10}$ was compared to Bayesian weighting when all 16,384 sweeps were utilized in the average. Thus it appears, and as expected, that the total number of sweeps that were averaged had a large impact on the RN.

From their study the authors concluded that AR and Bayesian weighting are equally effective when testing quiet and relaxed participants. They also found that when participants were behaviorally active, a strict AR level of ±10 μV can reduce noise (when using a fixed number of accepted sweeps of 4096) but at the cost of inefficiency (averaging will take much longer to obtain a sufficient number of sweeps).

A more recent study by Cone and Norrix (2015) obtained ABR thresholds in 40 adults with normal hearing who participated in a quiet condition (2000 total sweeps) and either a steady state or intermittent motor activity condition (3 minutes of averaging). The motor activity for the steady state condition consisted of participants reading aloud. For the intermittent condition, participants performed a motor task (humming, writing in the air, counting, or listing objects or names) every 30 seconds for 10 seconds during averaging. Presumably, if wave V thresholds are similar, the RN levels should be similar. If a threshold is elevated, the response was likely embedded in higher levels of RN.

Cone and Norrix recorded ABRs using traditional averaging with AR set at 20µV (± 10µV) and (KWA). The traditional averaging plus AR was implemented using the IHS. Kalman-weighted averaging was implemented using the Vivosonic Integrity v5.0 software. Kalman weighting, like Bayesian weighting, assigns greater weight to quiet sweeps and less weight to noisy sweeps during the averaging process (Hall, 2007; Kay, 1993). In addition, the Vivosonic employs an in-situ amplifier (Amplitrode) and Bluetooth technology which can further reduce unwanted noise in recordings (The Amplitrode, n.d.). The authors found that ABR thresholds in quiet physiological states were similar for traditional averaging plus AR and KWA. In comparison to the ABR thresholds obtained in quiet, in the steady state motor activity conditions mean ABR thresholds were elevated by 18.5 dB (SD of 13.7 dB) for traditional averaging plus AR and by 11 dB (SD of 9.7 dB) for KWA. Thus, there was a 7.5 dB threshold advantage for KWA compared to AR. In the intermittent motor noise condition, compared to the Quiet thresholds, mean ABR thresholds were elevated by 1.3 (SD of 8.6 dB) and 3.6 dB (SD of 10.3 dB) for traditional averaging with AR and KWA, respectively. The authors concluded that in their intermittent noise condition, the two noise reduction methods resulted in comparable ABR thresholds.

Prior research has contributed to our knowledge of techniques to reduce residual noise in electrophysiological recordings. It appears that the type of motor activity and frequency of motor activity are critical factors when comparing noise reduction methods. The length of time that averaging is performed and total number of sweeps that are averaged will influence the quality of the ABR recordings. Finally, an examination of the SDs of the RN obtained in prior studies can be quite high. This is likely a result of the variability in the noisiness or quietness of individual participants. That is, even in a quiet condition it is likely that some participants are noisier than others or have some intermittent motor activity. In active motor conditions, it is very unlikely that all participants will perform a motor activity in a similar manner or with the same degree of motor activity. Although not a critical factor when a single set of recorded sweeps is submitted to offline analyses, when ABR recordings are made sequentially and compared there is a high likelihood that the noise amplitudes are variable between the two recordings.

No study, to our knowledge, has directly examined RN values obtained using AR and weighted averaging using a predetermined averaging time. Sanchez and Gans (2006) used a set number of sweeps and found that an AR setting of ±10 μV resulted in a lower RN than did Bayesian weighting, however, averaging was not efficient as the mean number of rejected sweeps to attain their criterion of 4096 accepted sweeps, was 4,234 (SD of 2,220). Cone and Norrix (2015) specified a criterion averaging time in their study, however, they did not examine RN but determined wave V thresholds as a function of noise reduction strategy. Therefore, in this study, our aim was to examine RN levels as a function of averaging time when using traditional averaging plus AR and KWA. We compared RN measures obtained by two commonly used manufacturer-specific ABR devices: the IHS (traditional averaging and artifact rejection) and the Vivosonic Integrity (Kalman weighted averaging, the Amplitrode, and Bluetooth). Residual noise levels were compared for relaxed and active participant motor states using simultaneous recordings to ensure similar electrophysiological noise levels recorded by the two ABR devices. Specifically, we examined whether traditional averaging/AR and Kalman-weighted averaging/Amplitrode/Bluetooth technologies were successful in reducing RN to a low criterion level (effectiveness of the noise reduction strategy). We also examined the relative amount of time needed to get to the low noise criterion (efficiency of the noise reduction strategy).

# METHODS

*Participants*

Sixteen adults, 5 men and 11 women, between the ages of 21 and 38 years (M=25.9 years, SD=4.4 years) were recruited from the University of Arizona, following approval by the Social and Behavioral Sciences division of the Human Subjects Protection Program at the University of Arizona. Each participant signed an informed consent form, approved by an institutional review board. Otoscopic examinations were performed prior to ABR testing to ensure no visible patholgy in the ear canals that could preclude use of insert earphones.

*Stimuli and ABR Acquisition Parameters*

The Vivosonic Integrity System - V500 G1 System (VIVO) high- and low-pass filter roll-offs were 12 dB/octave and 24 dB/octave, respectively. High- and low-pass filter roll-offs were 6 dB/octave for the Intelligent Hearing Systems – Smart EP (IHS). The IHS amplifier had a gain of 100 K, while the VIVO gain was 7,500. Both devices had EEG filter settings of 100 – 1500 Hz. Electrode impedances were kept below 3 kΩ throughout the duration of the experiment.

Stimuli were used to initiate the recordings. Inaudible rarefaction clicks presented at -15 dB nHL at a rate of 37.7/s, through a VIVO insert earphone, were used for all ABR recordings.

*Procedure*

The forehead and earlobes, back and front, of each participant were cleaned with an alcohol wipe to remove oils from the skin and then were scrubbed with NuPrep, an abrasive gel. Six disposable adhesive electrode tabs were then placed: two on the forehead, one on the front of each earlobe and one on the back of each earlobe. Three electrodes were connected to the IHS ABR system and three electrodes were connected to the VIVO ABR system. The noninverting electrodes were placed on to the right and left of Fz. The inverting electrodes were always on the left earlobe (A1) and the ground electrodes were always on the right earlobe (A2). Half of the participants had front earlobes and left Fz montage for the VIVO recordings and back earlobes and right Fz montage for the IHS recordings.  The other eight participants had front earlobes and a left Fz montage for the IHS recordings and back earlobes and a right Fz montage for the VIVO recordings. (Figure 1).

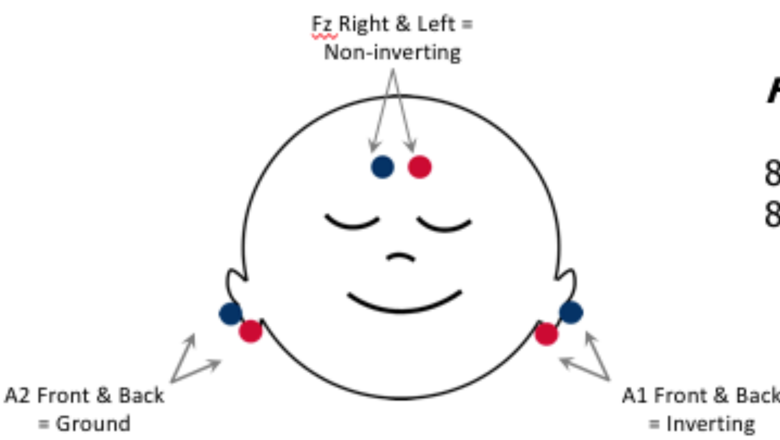



***Figure 1.*** There were two electrode montages:

8 participants had ● IHS and ● VIVO;
8 participants had ● IHS and ● VIVO.

Participants were seated in a reclined chair in a double-walled sound treated booth. Electroencephalogram (EEG) was recorded simultaneously using IHS and VIVO ABR devices using a sampling rate of 20,000 Hz.  Because the VIVO ABR system has an acquisition delay of about 9 seconds, the IHS recording was started 9 seconds after the VIVO.  Each participant was evaluated in 6 different conditions (Table 1).  There were two activity states (Relaxed and Active).

<u>Relaxed Conditions:</u>

Each adult participated in two relaxed conditions.  For both conditions, participants were seated in a reclined chair and encouraged to relax, keep eyes closed and try to sleep.  For relaxed state 1, artifact rejection (AR) was set to the highest level allowed by the IHS system – 100µV and the Kalman weighted averaging (KWA) was turned off on the VIVO system.  For relaxed state 2, AR was set to 10µV (±5 µV) and KWA was on.  The relaxed state 1 and 2 were counterbalanced across participants. Three averaging times were used for each condition: 1, 1.5, and 3 minutes.

Active Conditions:

Artifact rejection or KWA were used for all conditions. Each adult participated in four active conditions. For all conditions participants were seated in a reclined chair and were instructed to constantly or intermittently perform a motor activity during the ABR recordings. In the constant condition, they performed a motor activity during the duration of the averaging time. In the intermittent condition, they performed the motor activity as soon as averaging was initiated for a duration of 15 seconds. Following the 15 seconds of motor activity, participants were instructed to close eyes and return to a relaxed state. Each participant was lightly tapped as a cue to begin and end the motor activity. The motor activity recurred after 1 minute of averaging had transpired.

The motor activities consisted of playing a game on an iPad and opening and closing the mouth while nodding the head up and down. The iPad motor activity was performed intermittently in one condition and constantly in a comparable condition. AR was set to 10 μV and KWA was on. Three averaging times were used for these conditions (1.5, 3, and 6 minutes). The head nod/mouth open motor activity was performed in two intermittent conditions. In the first condition, AR was set to 10μV and KWA was on. Three averaging times were used (1.5, 3 and 6 minutes). In the second condition, AR was set to 20μV (±10 μV) and KWA was on. Only a 6-minute averaging time was used (refer back to Table 1). The order of the active conditions remained consistent across all participants as indicated in the table.

| Condition | Participant Activity | Parameters | Time (min.) |
|---|---|---|---|
| Relaxed State 1 | Relaxing. | IHS: AR = 100µV; VIVO: KWA = off | 1, 1.5, 3 |
| Relaxed State 2 | Relaxing. | IHS: AR = 10µV; VIVO: KWA = on | 1, 1.5, 3 |
| Intermittent Active 1 (iPad) | Intermittent iPad gaming (15 seconds each minute). | IHS: AR = 10µV; VIVO: KWA = on | 1.5, 3, 6 |
| Constant Active (iPad) | Playing a game on the iPad. | IHS: AR = 10µV; VIVO: KWA = on | 1.5, 3, 6 |
| Intermittent Active 2 (Mouth) | Intermittent head nodding with opening/closing mouth (15 seconds each minute). | IHS: AR = 10µV; VIVO: KWA = on | 1.5, 3, 6 |
| Intermittent Active 3 (Mouth) | Intermittent head nodding with opening/closing mouth (15 seconds each minute). | IHS: AR = 20µV; VIVO: KWA = on | 6 |

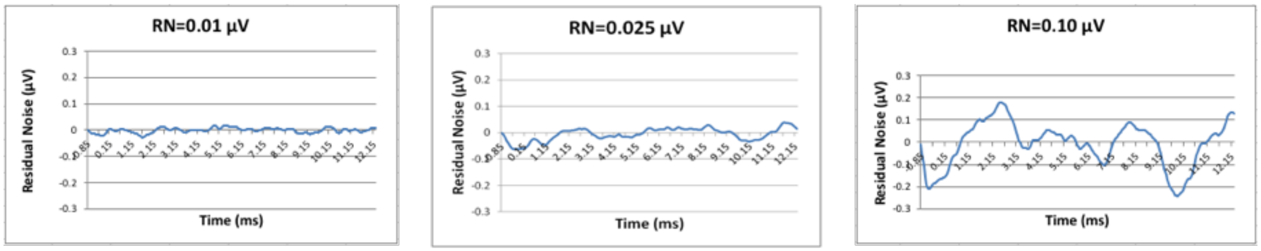
***Table 1.*** Description of abbreviations: AR = artifact rejection; KWA= Kalman weighted averaging. Relaxed States 1 and 2 were counterbalanced. Total number of possible sweeps per time epoch: 1 min. ≈ 2200, 1.5 min. ≈ 3400, 3 min. ≈ 6800, and 6 min. ≈ 13,500. When participants were not engaging in motor activity during "Intermittent" active conditions, they were relaxing.

*Dependent Measure and Data Analyses*

Our dependent measure was residual noise level. For both the IHS and the VIVO measures, RN levels were derived from calculating the standard deviation of the digitized data for each average recording (in ≈ 0.05ms increments) from -0.85 ms to 12.79 ms. The data were submitted to repeated Analyses of Variance (ANOVA) using the Statistical Package for Social Sciences Versions 25.0.

Effectiveness of the noise reduction strategy in reducing RN was examined by determining if the mean RN was reduced to a RN level criterion of 0.025 µV. This criterion was chosen as we believed that if RN was 0.025 µV or less, a wave V response of the ABR, if present, would be evident in the majority of averaged waveforms. Figure 2 provides examples of RN calculated in this manner. Our second goal was to determine the efficiency of the noise reduction method. The relative length of averaging time needed to get to the criterion RN level was examined as a measure of efficiency.

***Figure 2.*** Visual representation of variance of the waveforms and the corresponding RN levels derived by finding the standard deviation of all the data points from -0.85 – 12.79 ms.



# RESULTS

## *Relaxed State Condition*

Figure 3 shows the mean RN levels obtained in the relaxed state condition. In this condition, KWA was more effective in reducing RN than were the other methods (i.e., traditional averaging or tradition averaging with AR). However, even with the use of KWA, 3 minutes of averaging was needed for the majority of participants (11/16) to reach the low RN criterion. Note that in this "relaxed state" condition, participants were instructed to close their eyes and relax, to try to fall asleep, to keep still and to be in a completely relaxed state. However, there was a large degree of variability in how relaxed our participants were. Because simultaneous recordings were used, an examination of the percentage of rejected trials using the IHS with an artifact rejection setting of 10 µV provides an indication of the activity level of our participants during the recordings. Table 2 shows that while some participants had very few artifact rejections, others had a large number of rejected trials and therefore were not likely in a physiologically restful state during the recordings.

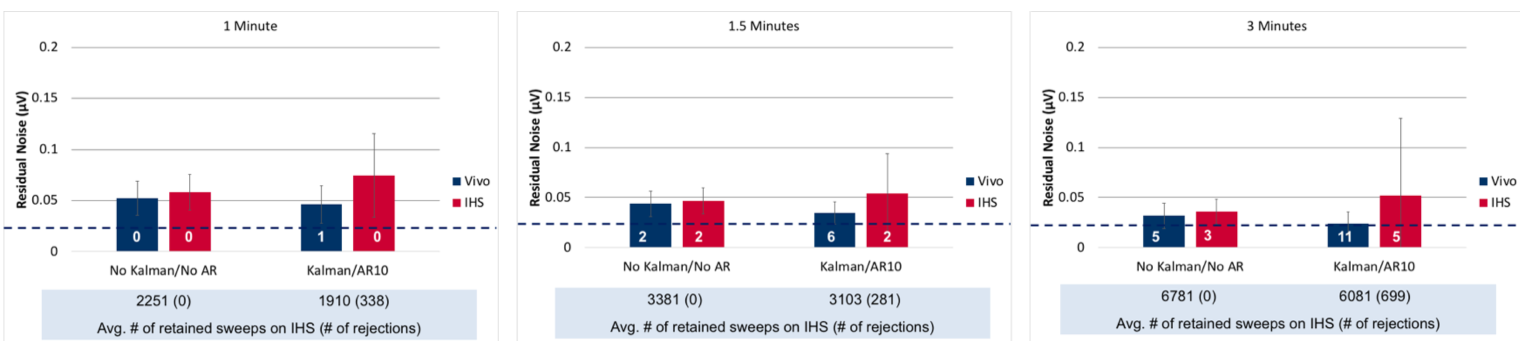



**Figure 3.** Average residual noise (RN) levels, $\pm$1 standard deviation, obtained in the Relaxed State Condition for 1, 1.5, and 3 minutes of averaging. The dashed line indicates a RN criterion of 0.025 µV. The number of participants, out of 16, who met the criterion is indicated within each bar on the graph.

**Table 2.** Percent of rejected trials in the relaxed condition with an AR setting of 10 µV.

| | 1 minute | 1.5 minutes | 3 minutes |
|---|---|---|---|
| Range (%) | 0-57 | 0-44 | 0-76 |
| Mean (%) | 15 | 8 | 10 |
| Standard Deviation (%) | 17 | 12 | 19 |

A three-way ANOVA, using Greenhouse-Geisser corrections, with Time (1, 1.5, 3 minutes), Device (IHS, VIVO), and NR strategy (on, off) as within subject variables was performed to examine the differences between RN levels. The results (Table 3) revealed a main effect of device [$F$ (1, 15) = 10.91, $p$=0.005] with the VIVO recordings having a lower average RN ($M$=0.039 µV) than the IHS ($M$=0.053 µV). Also significant was the main effect of time, [$F$ (1.3, 19) = 17.46, $p$<0.001]. Paired comparisons, adjusted for multiple comparisons, showed that the RN was significantly lower with 3 minutes of averaging ($M$=0.036 µV) compared to 1.5 and 1 minutes of averaging. In addition, the RN for 1.5 minutes of averaging was significantly lower ($M$=0.045 µV) than for 1 minute of averaging ($M$=0.058 µV). This was expected because with averaging, noise is reduced by the square root of the number of sweeps and therefore the greater the number of sweeps, the lower the RN. In this relaxed condition, there was no significant effect of having the noise reduction strategy on versus off and no significant interactions between any of the variables.

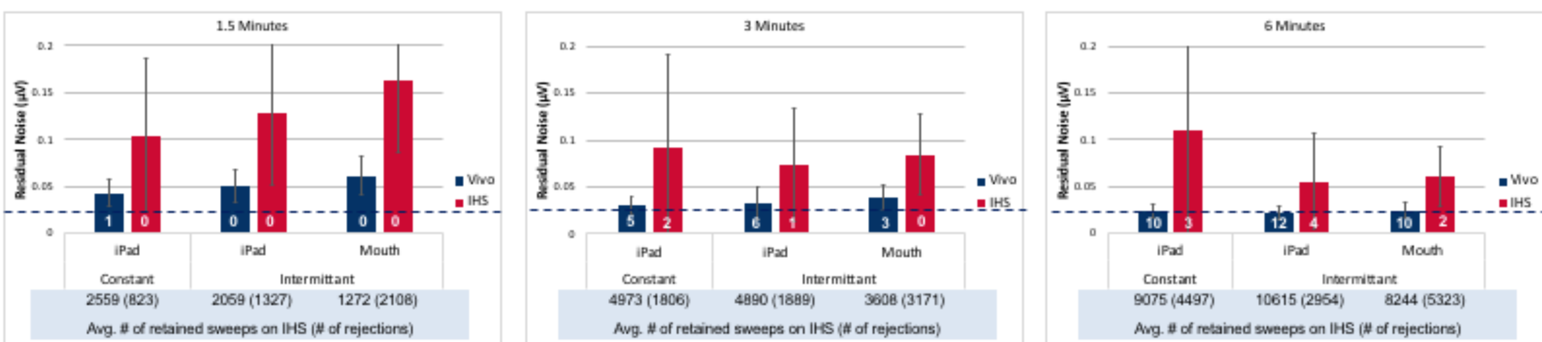
**Table 3.** Three-way ANOVA results for relaxed conditions

| Source | *df* | *F* | *p* |
|---|---|---|---|
| NR Strategy (on/off) | 1 | 0.16 | 0.698 |
| Time | 1.3 | 17.46 | <0.001 * |
| Device | 1 | 10.91 | 0.005 * |
| NR Strategy (on/off) x Time | 1.6 | 0.79 | 0.44 |
| NR Strategy (on/off) x Device | 1 | 3.58 | 0.08 |
| Time x Device | 1.6 | 0.71 | 0.47 |
| NR Strategy (on/off) x Time x Device | 1.4 | 0.15 | 0.78 |

*=significant at $p < 0.05$

*Active State Condition*

Figure 4 shows the mean RN levels obtained in the active state conditions. In the active states, KWA was more effective in reducing RN than was traditional averaging with AR. By using KWA, the RN criterion level was met, on average, by 6 minutes for each type of motor activity. This criterion RN level was not met using traditional averaging with artifact rejection set at 10 µV. Use of AR was very ineffective for most of the participants, although as seen in Figure 4, a low RN level was obtained for a few participants particularly in the 6-minute averaging condition. Conversely, although KWA was more effective than AR, a few participants did not reach the low RN criterion using KWA even after 6 minutes of averaging. Individual variability in terms of the amount and amplitude of motor activity was high. Table 4 shows the percent of rejected trials in the active conditions with an AR setting of 10 µV. As can be seen in this table, the percentage of rejected sweeps ranged from 0 to 90% for the iPad Constant condition and 0 to 76% for the iPad Intermittent condition. Thus, while some participants were rather relaxed while performing the iPad game, others performed the activity with greater motor




involvement.

**Figure 4.** Mean residual noise (RN) levels, ±1 standard deviation, obtained in the Active State Conditions for 1.5 (A), 3 (B), and 6 minutes (C) of averaging. The dashed line indicates a RN criterion of 0.025 µV. The number of participants, out of 16, who met the criterion is indicated within each bar on the graph.

**Table 4.** Percent of rejected trials in the active conditions with an AR setting of 10 µV.

| | iPad Constant | | | iPad Intermittent | | | Mouth Intermittent | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1.5 minutes | 3 minutes | 6 minutes | 1.5 minutes | 3 minutes | 6 minutes | 1.5 minutes | 3 minutes | 6 minutes |
| Range (%) | 0-81 | 0-78 | 0-90 | 2-64 | 1-76 | 0-60 | 39-78 | 27-79 | 26-63 |
| Mean (%) | 24 | 27 | 33 | 39 | 28 | 22 | 62 | 47 | 39 |
| Standard Deviation (%) | 22 | 22 | 27 | 19 | 21 | 18 | 12 | 13 | 12 |

A three-way ANOVA, using Greehouse-Geisser corrections, with Time (1.5, 3, 6 minutes), Device (IHS, VIVO), and Activity (intermittent iPad, intermittent mouth, constant iPad) as within subject variables was performed to examine the differences between RN levels. The results (Table 5) revealed a main effect of Device [$F$ (1,15) = 27.63, $p$<0.001] with the VIVO recordings having a lower average RN ($M$=0.038 µV) than the IHS ($M$=0.094 µV). Also significant was the main effect of Time, [$F$ (1.2, 18.0) = 50.3, $p$<0.001]. Paired comparisons, adjusted for multiple comparisons, showed that, as expected, the RN was significantly lower with 6 minutes of averaging ($M$=0.040 µV) compared to 3 and 1.5 minutes of averaging. The RN for 3 minutes of averaging ($M$=0.057 µV) was also significantly lower than for 1.5 minutes of averaging ($M$=0.101 µV).

Additionally, there were significant two-way interactions between Time and Device [$F$ (1.5, 22.4) = 10.12, $p$=0.002], and between Activity and Time [$F$ (2.7, 41.0) = 15.91, $p$=<0.001]. A significant 3-way interaction was also obtained between Activity, Time, and Device [$F$ (2.4, 36.0) = 9.65, $p$<0.001].

**Table 5.** Three-way ANOVA results for active conditions.

| Source | *df* | *F* | *p* |
|---|---|---|---|
| Activity | 1.4 | 0.55 | 0.524 |
| Time | 1.2 | 50.33 | <0.001 * |
| Device | 1 | 27.63 | <0.001 * |
| Activity x Time | 2.7 | 15.91 | <0.001 * |
| Activity x Device | 1.3 | 0.484 | 0.542 |
| Time x Device | 1.5 | 10.12 | 0.002 * |
| Activity x Time x Device | 2.4 | 9.65 | <0.001 * |

*=significant at $p < 0.05$

Figure 5 graphically represents the 3-way interaction. As can be seen in this figure, RN levels for VIVO as a function of averaging time did not vary by noise type (iPad vs. mouth/head movement) or by frequency (intermittent vs. constant), whereas the duration and amplitude of motor activity had a large impact on RN level as a function of time for the IHS.

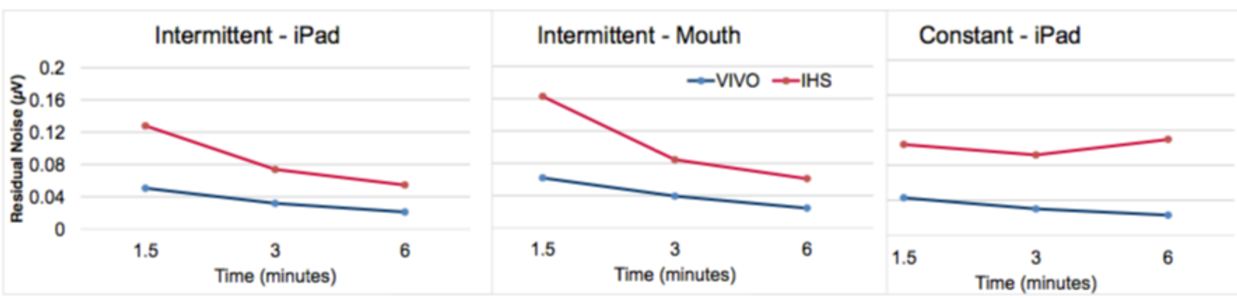



**Figure 5.** Residual noise as a function of Time, Device, and Activity.

*Artifact Rejection Setting*

We expected a high percentage of rejections in our active conditions, particularly with an AR setting of 10 µV. This would result in a reduced number of sweeps in the final average. In order to obtain a greater number of sweeps in the average the AR setting was increased from 10 to 20 µV for the Intermittent Mouth Condition in the 6-minute averaging interval. Figure 6 shows the RN when using an AR setting of 10 µV (from Figure 4C) compared to the same condition using a 20 µV AR setting.

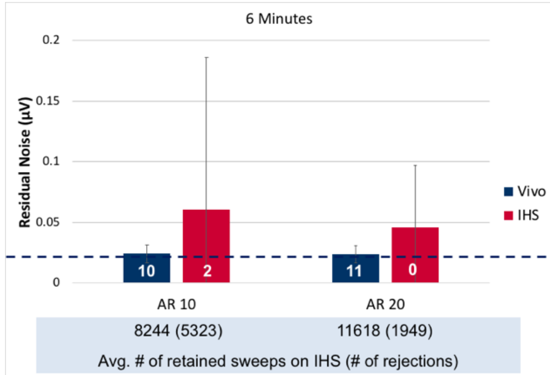



**Figure 6.** Mean residual noise (RN) levels, +1 standard deviation, obtained in the Intermittent Mouth Conditions for 10 µV and 20 µV. The dashed line indicates a RN criterion of 0.025 µV. The number of participants, out of 16, who met the criterion is indicated within each bar on the graph.

A two-way ANOVA, using Greenhouse-Geisser corrections, with Device (IHS, VIVO) and AR Setting (10 µV, 20 µV) as within subject variables was performed to examine the differences between RN levels as a function of artifact rejection setting. The results (Table 6) showed a significant main effect of Device [$F$ (1, 15) = 89.12, $p$=<0.001] with the VIVO having a lower overall RN ($M$=0.024 µV) than the IHS ($M$=0.053 µV). The AR setting as well as the interaction between Device and AR setting were not significant.

**Table 6.** Two-way ANOVA comparing 2 artifact rejection settings.

| Source | *df* | *F* | *p* |
|---|---|---|---|
| AR Setting | 1 | 3.93 | 0.067 |
| Device | 1 | 89.12 | <0.001* |
| AR Setting x Device | 1 | 3.15 | 0.096 |

*=significant at $p < 0.05$

Increasing the AR level to 20 µV did result in fewer rejected sweeps. Table 7 shows that on average, 39% of the sweeps were rejected using a setting of 10 µV while only 14% were rejected using a setting of 20 µV. Although not significant, there was a trend for a lower RN with an AR setting of 20 µV ($M$=0.035 µV) compared to when using a setting of 10 µV

(*M*=0.042 μV), presumably due to a greater number of sweeps included in the average. Despite the greater number of sweeps and 6 minutes of averaging, no participant had a RN level of 0.025μV or less.

**Table 7.** Percentage rejected trials in the active mouth condition, 6 minutes of averaging, with an AR setting of 10 μV compared to 20 μV.

| | 10 μV | 20 μV |
|---|---|---|
| Range (%) | 26-63 | 2-29 |
| Mean (%) | 39 | 14 |
| Standard Deviation (%) | 12 | 7 |

# DISCUSSION

Artifact rejection (AR) and weighted averaging are techniques commonly used to reduce noise during ABR testing (Don & Elberling, 1994; Elberling & Wahlgreen, 1985). The current study was designed to test the effectiveness and efficiency of the IHS (using artifact rejection) and VIVO (using Kalman weighted averaging, the Amplitrode, and Bluetooth) ABR systems in reducing residual noise (RN). Adults were tested in a quiet, relaxed state and also during periods of induced motor movement. Measures from each system were simultaneously recorded to eliminate any differences in noise levels due to sequential recording from each system individually, thus any noise present would be at similar levels for both systems. A residual noise level of 0.025µV was the criterion for effectiveness. Averaging time to examine efficiency ranged from 1-3 minutes for the Relaxed Conditions and 1.5-6 minutes for the Active Conditions.

In the relaxed state conditions, utilization of KWA was overall more effective at reaching low residual noise levels when compared to AR. This finding is in contrast to Sanchez and Gans (2006), who found no difference in residual noise levels when using weighted averaging compared to AR. This difference in findings could be due to the differences across studies in how relaxed participants were. While it is not clear how physiologically "quiet" the participants were in the Sanchez and Gans study, in the present study there was a large amount of variability in the number of sweeps that were considered too noisy and were rejected in the relaxed state; thus, while some participants were in a physiologically quiet state, others were not. Differences in physiologic noise during "relaxed" conditions may result in different outcomes between studies.

In the active state conditions, utilization of KWA was overall more effective at reaching low residual noise levels when compared to AR. This finding is consistent with the Don and

Elberling (1994) study. In contrast, Sanchez and Gans (2006) found an AR of ±10 μV to be more effective at reducing noise than weighted averaging during active conditions. This difference could be due to several factors. First, there could be a difference between the amount and frequency of the motor activity of the participants between the two studies. Additionally, there was a difference in the number of sweeps averaged between the studies. In the Sanchez and Gans study, averaging occurred for the first 4096 sweeps when using weighted averaging and for the first 4096 accepted sweeps when using AR. In contrast, the present study used a set amount of averaging time rather than a set number of sweeps in order to simulate clinically relevant time intervals. Therefore, the number of sweeps contributing to the average when using AR was smaller than that of the number of sweeps for KWA.

In the current study, RN levels were lower when using KWA compared to AR regardless of the frequency of the motor activity (constant versus intermittent). Although no study to our knowledge has examined residual noise as a function of the frequency of the motor activity, Cone and Norrix (2015) calculated ABR wave V threshold differences, for quiet versus active conditions, when using AR compared to KWA. As a high RN in the average will make it difficult to detect a threshold wave V response, the assumption is that the smaller the threshold difference between the quiet and active conditions, the lower the RN in the active condition. Cone and Norrix (2015) found that the threshold differences were similar for AR and KWA averaging when subjects were asked to produce intermittent motor activity. However, in the steady state noise condition there was a 7.5 dB advantage for KWA compared to AR. Differences between studies; however, may be a result of differences in the amount of noise produced by the participants or differences in the relative noise amplitude during the intermittent activity.

In terms of efficiency, even when using KWA, neither 1 nor 1.5 minutes of averaging were sufficient for producing an average RN level that met the criterion. Using KWA, it wasn't until 3 minutes of averaging that the majority of participants were able to meet the criterion RN level. Similarly, for the active state conditions it wasn't until 6 minutes of averaging that the criterion RN level was met. Averaging for 1.5 and 3 minutes was not sufficient to generate a low RN level to meet the criterion.

*Limitations*

One limitation of the present study is that it is difficult to compare this research to what is used clinically. Typically, ABR recordings are used for determining hearing sensitivity in infants, but the participants used in the present study were adults. It is possible that sleeping infants are much quieter than our relaxed adults during ABR recordings. Additionally, the present study did not look at RN when a stimulus was present, and it is possible that the presence of a stimulus could impact residual noise. This being said, future studies should examine residual noise with varying stimuli present.

*Conclusions*

The results from the current study provide insight into the effectiveness and efficiency of noise reduction techniques during ABR recordings under relaxed and active participant motor states. The noise reduction techniques examined were Kalman weighted averaging, the Amplitrode, and Bluetooth (implemented on the VIVO) and AR (implemented on the IHS). The following conclusions are made:

1. VIVO was more effective than the IHS. With use of KWA, the Amplitrode and Bluetooth technology, a lower RN level was obtained for all conditions compared to the IHS using traditional averaging or traditional averaging with artifact rejection.

2. In the relaxed state, the criterion residual noise level was only met by the VIVO using KWA, the Amplitrode and Bluetooth. Additionally, using the VIVO it took 3 minutes of averaging in order for the majority of participants to meet the criterion residual noise level.
3. In the active state, the criterion residual noise level was only met by the VIVO using KWA, the Amplitrode and Bluetooth. Six minutes of averaging was needed for the residual noise criterion to be met by the VIVO.
4. While increasing the AR from 10 μV to 20 μV during the 6-minute intermittent mouth condition resulted in fewer rejected sweeps, the average RN still did not reach the criterion RN level.
5. While it might be possible to use KWA on an active infant, it may not be clinically feasible. Depending on the activity level of the infant, it could take over 6 minutes to obtain a response for each frequency and even then, it is not guaranteed that the RN would be low enough to accurately determine if a response was or was not present.

# REFERENCES


Cone, B., & Norrix, L. W. (2015). Measuring the advantage of Kalman-weighted averaging for auditory brainstem response hearing evaluation in infants. *American journal of audiology*, *24*(2), 153-168.

Don, M., & Elberling, C. (1994). Evaluating residual background noise in human auditory brain-stem responses. *The Journal of the Acoustical Society of America*, *96*(5), 2746-2757.

Elberling, C., & Don, M. (1984). Quality estimation of averaged auditory brainstem responses. *Scandinavian audiology*, *13*(3), 187-197.

Elberling, C., & Wahlgreen, O. (1985). Estimation of auditory brainstem response, ABR, by means of Bayesian inference. *Scandinavian Audiology*, *14*(2), 89-96.

Hall, J. W. (2007). *New handbook of auditory evoked responses* (Vol. 1). Boston: Pearson.

Kay SM. (1993). *Fundamentals of statistical signal processing.* Eaglewood Cliffs, NU: Prentice Hall.

Lightfoot, G., & Stevens, J. (2014). Effects of artefact rejection and Bayesian weighted averaging on the efficiency of recording the newborn ABR. *Ear and hearing*, *35*(2), 213-220.

Maybeck PS. (1979). *Stochastic models, estimation and control.* New York, NY: Academic Press.

Newborn Hearing Screening and Assessment (2013). Guidelines for the early audiological assessment and management of babies referred from the newborn hearing screening programme, version 3.1. Retrieved from htttp://www.thebsa.org/uk/wp-content/uploads/2014/08/NHSP_NeonateAssess_2014.pdf

Picton, T. W. (2011). *Human auditory evoked potentials*. Plural Publishing.

Sanchez, J. T., & Gans, D. (2006). Effects of artifact rejection and Bayesian weighting on the auditory brainstem response during quiet and active behavioral conditions. *American journal of audiology*, *15*(2), 154-163.

"The Amplitrode." (n.d.) describes the Amplitrode used by Vivosonic Inc. Retrieved from www.vivosonic.com/amplitrode.html

Upp, Daniel C. (2002). *U.S. Patent No. 6,463,111*. Shelton, CT: U.S. Patent and Trademark Office.